Self-calibrating, environment-adaptive physical neural networks (PNNs) for robust and energy-efficient AI

Angshuman Mukherjee, Enterprise Architect, CTO Team, Global Client Growth Office

Contents

1. Executive summary

Artificial intelligence (AI) is rapidly increasing the world's demand for computational power, energy and infrastructure. Traditional digital processors are approaching the practical limits of semiconductor scaling, creating a need for fundamentally new computing architectures that deliver significantly higher efficiency. **Physical neural networks** (PNNs) — implemented using photonic, memristive and analog in-memory computing technologies — represent one of the most promising alternatives, offering the potential for orders-of-magnitude improvements in energy efficiency compared with conventional digital accelerators.

Despite their promise, the widespread deployment of PNNs faces a critical obstacle: **physical non-ideality**. Real-world hardware is affected by fabrication variability, thermal drift, supply-voltage fluctuations, noise, and device aging. As a result, systems calibrated under laboratory conditions gradually lose accuracy once deployed. Existing solutions rely on periodic recalibration, which consumes energy, increases operational cost and undermines many of the efficiency benefits that make PNNs attractive in the first place.

This paper presents a novel research vision for **self-calibrating, environment-adaptive physical neural networks** that continuously monitor and correct their own physical state without interrupting normal operations. The central hypothesis is that physical drift should not be treated merely as an engineering problem to be tolerated or periodically corrected, but rather as a measurable and learnable signal that can be managed through intelligent control mechanisms.

The proposed architecture combines **a primary physical neural network** with **an embedded sensing subsystem** and **a lightweight meta-controller**. Sensors continuously monitor key physical parameters such as temperature, conductance drift, optical power and supply voltage. The meta-controller uses these signals to infer the current health and state of the physical substrate and applies targeted corrections in real time. Rather than recalibrating an entire network, the controller selectively adjusts only the most influential parameters, dramatically reducing energy overhead while preserving accuracy.

A major innovation is the meta-controller's **closed-loop learning framework**, which integrates sensing, adaptation and inference into a continuously operating feedback system. Advanced machine learning techniques, including meta-learning, continual learning and reinforcement, enable the meta-controller to adapt rapidly to new environmental conditions, hardware aging effects and unforeseen operating scenarios. The framework is designed to maintain system stability while balancing three competing objectives: task accuracy, energy efficiency and long-term hardware reliability.

This paper proposes a comprehensive **three-phase validation program**. Phase 1 focuses on simulation and algorithm development using realistic models of photonic and memristive hardware. Phase 2 transitions to prototype implementations on field-programmable gate array (FPGA)-emulated and silicon photonic platforms. Phase 3 evaluates performance under real-world conditions, including wide temperature variations, long-duration operation and power-supply disturbances. Success is measured through metrics such as sustained accuracy, adaptation latency, energy-overhead reduction and deployment-lifetime extension.

The architecture aims **to reduce calibration overhead** by more than 80%, maintain over 95% of **baseline accuracy under challenging environmental conditions**, and **extend practical hardware deployment lifetimes** by three to five times those of conventional static calibration approaches. Furthermore, the solution seeks to **preserve most of the energy-efficiency advantages** that make PNNs attractive, enabling practical deployment in edge devices where power, cost and maintenance constraints are severe.

Beyond its technical contributions, the proposed three-phase program has significant **economic and societal implications**. Self-calibrating PNNs could enable a new generation of ultra-efficient AI systems for industrial IoT, automotive and aerospace platforms, telecommunications infrastructure, healthcare wearables, and low-cost edge devices in emerging markets. By reducing both computational energy consumption and hardware replacement frequency, the technology would also support broader sustainability objectives, including lower carbon emissions and reduced electronic waste.

2. Abstract

Physical neural networks (PNNs), implemented on photonic, memristive and analog in-memory substrates, represent a defining frontier in energy-efficient artificial intelligence. These systems offer orders-of-magnitude improvements in computational energy efficiency over conventional digital accelerators. However, their practical deployment is severely constrained by **intrinsic physical non-idealities**: fabrication variability, thermal drift, voltage fluctuations and device aging. In practice, this means a PNN calibrated in the lab drifts out of specification once deployed in the field, and today's fix — periodic recalibration — consumes the very energy savings PNNs are meant to deliver.

Existing calibration strategies impose unacceptable overhead and fail to generalize across real-world deployment environments. This paper presents a novel research vision for self-calibrating, environment-adaptive PNNs that integrate embedded physical sensing, meta-learning control and closed-loop training frameworks.

The proposed architecture couples a primary physical computation layer with a lightweight meta-controller that **continuously senses and adapts** to the physical substrate state — dynamically correcting for drift and noise without interrupting inference.

Key projected outcomes of the approach proposed in this paper include:

- Reduction in calibration overhead by up to 80%
- Extension of practical deployment lifetimes by a factor of three to five
- Enablement of practical AI inference on low-cost, resource-constrained edge devices
- Democratizing the use of AI in emerging markets by making AI hardware far less costly
- Applying AI to sustainability-critical applications

3. Introduction

The computational demands of modern artificial intelligence (AI) are increasing at a rate that outpaces improvements in digital semiconductor efficiency. While historical Moore's Law scaling delivered consistent performance gains, the physical limits of silicon Complementary Metal-Oxide-Semiconductor (CMOS) technology now impose diminishing returns. This trajectory has catalyzed intense interest in alternative computing substrates that can perform AI inference and training with fundamentally lower energy expenditure.

One such promising alternative is a physical neural network, or PNN. A PNN does not simulate a neural network in software; it builds mathematics directly into a physical process. Where a conventional AI chip represents a number as a pattern of digital bits and steps through billions of transistor-switching events to multiply and add those numbers, a photonic PNN can perform the equivalent operation the instant light passes through a shaped waveguide junction, and a memristive PNN can perform it the instant current flows through a resistive crossbar. This is the source of the dramatic energy savings PNNs promise, since no energy is spent on switching digital logic gates merely to emulate a calculation the physical substrate can perform on its own.

Physical neural networks (PNNs) have emerged as a compelling class of computing systems that harness the intrinsic physics of their substrates — light propagation, ionic drift and analog charge — to implement neural network computations directly in hardware. Rather than emulating neural computations digitally, PNNs leverage:

- Photonic neural networks to encode information in optical signals and exploit the ultrahigh bandwidth and near-zero propagation loss of waveguide-coupled components
- Memristive crossbar arrays to exploit nonlinear ionic transport to implement dense analog matrix multiplications, achieving energy efficiencies exceeding 10 TOPS/W (Yao et al. 2020)
- Reservoir computing systems to exploit the transient dynamics of complex physical systems — spintronic oscillators, mechanical structures — to perform high-dimensional nonlinear transformations

Despite these compelling capabilities, the translation of PNNs from controlled laboratory demonstrations to reliable real-world deployment remains hindered by a fundamental challenge: physical non-ideality. Real hardware components deviate from ideal specifications because of fabrication imperfections, thermal sensitivity, device aging and supply voltage fluctuations.

The problem behind this is that a PNN is calibrated once, under controlled laboratory conditions, to perform a specific computation correctly. From the moment it is deployed, however, the physical substrate it depends on keeps changing: chips warm up and cool down over the course of a day, individual devices wear differently as they age, and supply voltages fluctuate as other parts of a system draw power. Every one of these physical shifts nudges the PNN's output away from its calibrated answer. The only remedies available at present are to either recalibrate on a fixed schedule regardless of need, which consumes the energy budget the PNN was meant to save, or to accept the resulting accuracy loss, which is unworkable for any application that depends on the AI's output being correct. Neither option allows PNNs to move out of the lab and into everyday products.

Central hypothesis

A PNN that continuously senses and adapts to its own physical state can achieve higher levels of robustness, lower calibration overhead and greater efficiency than statically calibrated systems.

This would required building a small companion system alongside the PNN that constantly measures physical drift as happens via embedded temperature, voltage and device-state sensors, and feeds that reading into a lightweight controller trained to compute the correction needed to keep the PNN's output on target. Because this correction runs continuously and automatically in the background, the PNN never needs to be taken offline for scheduled recalibration and never accumulates the drift-related errors that constrain today's deployments.

This paper proposes a fundamentally new architecture for physical neural networks: one that is self-calibrating and environment-adaptive.

4. Technical background and state of the art

4.1 Physical neural networks: Substrate taxonomy

Physical neural networks encompass a diverse family of computing paradigms united by the principle that neural network computations are implemented through the physical dynamics of a substrate, rather than through digital logic.
Figure 1 illustrates the three primary implementation families: photonic neural networks, memristive neural networks and reservoir computing.

Physical neural networks (PNNs)

Photonic neutral networks	Memristive neutral networks	Reservoir computing
Architecture: Mach–Zehnder Interferometer (MZI) mesh on silicon photonic chips	**Architecture:** Crossbar arrays of memristive devices	**Architecture:** Fixed physical reservoir + Linear readout
Operation: Speed-of-light matrix multiplication	**Operation:** Ohm's Law dot products in hardware	**Operation:** Nonlinear transient dynamics as compute
Advantage: Wavelength-division multiplexing	**Advantage:** Dense analog weight storage	**Advantage:** No physical-layer weight training needed
Non-ideality: Thermal phase drift in phase-shifters	**Non-ideality:** Conductance drift over time	**Non-ideality:** Reservoir state drift
Efficiency: ~10–100 TOPS/W (published)	**Efficiency:** >10 TOPS/W (Yao et al. 2020)	**Efficiency:** Highly task-dependent

Common non-ideality challenges:
Fabrication variance, thermal drift, device aging, supply voltage noise

Figure 1: PNN substrate taxonomy: the three primary physical substrate families, their architectural principles, key advantages and characteristic non-ideality challenges

4.1.1 Photonic neural networks

Photonic neural networks exploit the interference and diffraction properties of coherent light to implement linear transformations, with nonlinear activation functions implemented in optical or electronic domains. The seminal work of Shen et al. (2017) demonstrated a fully optical neural network using a Mach–Zehnder interferometer (MZI) mesh on a silicon photonic chip. Key advantages include operation at the speed of light, inherent parallelism through wavelength-division multiplexing, and near-zero static power dissipation of passive components.

Key non-ideality: Thermal sensitivity of phase-shifters. Ambient temperature changes cause phase errors that accumulate across large MZI arrays, degrading the fidelity of matrix operations. This is particularly challenging in edge deployment environments with uncontrolled thermal conditions.

4.1.2 Memristive and in-memory computing

Memristive devices store analog weight values as conductance states and implement dot products through Kirchhoff's current law in crossbar arrays. Yao et al. (2020) demonstrated fully hardware-implemented memristive convolutional neural networks. Sebastian et al. (2020) reviewed memory device technologies for in-memory computing, encompassing phase-change memory (PCM), resistive RAM (RRAM), and ferroelectric devices.

Key non-ideality: Conductance drift. Both PCM and RRAM devices exhibit time-dependent changes in their programmed conductance values following deployment. Joshi et al. (2020) characterized this impact empirically, showing inference accuracy degradation from over 90% to below 60% over periods of hours without compensation — a critical barrier to practical deployment.

4.1.3 Reservoir computing and physical dynamics

Reservoir computing exploits the high-dimensional nonlinear transient dynamics of a fixed physical system as a computational resource, requiring only the training of a linear readout layer. Tanaka et al. (2019) provide a comprehensive review of physical reservoir computing implementations. The fixed nature of the physical reservoir is both an advantage — no weight training is required in the physical layer — and a limitation in terms of adaptability.

4.2 Non-ideality: The fundamental challenge

Physical non-ideality in PNNs manifests at multiple levels simultaneously. At the device level, individual components deviate from designed specifications because of nanoscale fabrication variance. At the circuit level, crosstalk, parasitic coupling and power-supply noise introduce correlated errors. At the system level, thermal gradients across a chip create spatially varying operating conditions.

Existing mitigation strategies fall into three categories: hardware-centric (improved device engineering), software-centric (noise-aware training), and systems-centric (periodic recalibration), none of which achieves the combination of low-cost, continuous adaptation and high accuracy that real-world deployment demands.

Non-ideality type	Primary substrate	Mechanism	Impact	Key reference
Thermal phase drift	Photonic	Temperature changes shift phase-shifter operating point	Accumulated matrix error	Shen et al. 2017
Conductance drift	Memristive	Time-dependent change in programmed resistance	Weight value corruption	Joshi et al. 2020
Write noise	Memristive	Variability in conductance achieved by write pulse	Weight precision loss	Yao et al. 2020
Read noise	Memristive	Cycle-to-cycle resistance measurement variation	Inference variance	Sebastian et al. 2020
Fabrication variance	All	Nanoscale manufacturing imperfections	Systematic offset errors	Miller 2015
Supply voltage noise	All	Power rail fluctuations	Operating point shifts	Nandakumar et al. 2018

Table 1: Different types of physical non-realities in PNNs

4.3 Meta-learning and adaptive control

Meta-learning ("learning to learn") provides algorithmic foundations for the adaptive control systems proposed in this paper. Finn, Abbeel, and Levine (2017) introduced model-agnostic meta-learning (MAML), enabling models to adapt rapidly to new tasks from few examples. This approach is well-suited to the PNN adaptation problem because physical drift scenarios share structural similarity, allowing a meta-trained controller to generalize rapidly to new drift states.

Reinforcement learning (RL) provides complementary tools for learning control policies in non-stationary environments. Online learning algorithms that update model parameters from streaming data without storing historical observations are particularly relevant to the low-memory constraint of edge-deployed meta-controllers.

4.4 Continual learning and non-stationary environments

Continual learning addresses the challenge of learning from a non-stationary data distribution without catastrophically forgetting previously acquired knowledge. Kirkpatrick et al. (2017) introduced elastic weight consolidation (EWC), which regularizes weight updates to protect important parameters from overwriting. In the context of PNN adaptation, the non-stationarity arises from the evolving physical state of the substrate, a novel variant of the continual learning problem.

4.5 Stochastic resonance and constructive noise

A provocative aspect of PNN operation is the possibility that noise can be constructively exploited rather than merely suppressed. Stochastic resonance (SR) is a well-established phenomenon in which adding an optimal level of noise to a subthreshold signal paradoxically improves signal detection and information transmission (McDonnell and Abbott 2009). This suggests that the noise environment of a PNN, rather than being merely tolerated, could be actively managed to enhance learning and inference performance — a paradigm shift that this research will specifically investigate.

5. Problem statement and research gaps

5.1 Conceptual problem definition

A physical neural network is designed to implement a target computational function under ideal conditions. In real-world deployment, the actual computation diverges from this ideal due to a combination of time-dependent physical changes — device aging and drift — and environment-dependent changes — temperature variation and supply voltage fluctuations. The goal of calibration is to track and compensate for this divergence continuously, maintaining the accuracy of the target function across the full deployment lifetime of the device.

A concrete illustration makes this tangible: Consider a memristive AI accelerator calibrated on a factory bench at a stable 25°C. Installed the same week inside an outdoor industrial sensor, its crossbar array now experiences a daily thermal swing of 40°C or more, on top of the slow conductance drift that memristive devices exhibit simply from being used. Within days, the chip's inference accuracy can fall from over 90% to well below the threshold needed for a safety-relevant decision — not because anything is broken, but because the physical assumptions built into its one-time calibration no longer hold. Multiply this scenario across millions of deployed edge devices and it becomes clear why physical non-ideality, more than raw computational capability, is the binding constraint on where PNNs can be used today.

Static calibration addresses the divergence at a single point in time under fixed conditions. This approach fails as devices age (drift), as operating temperatures change (thermal variation) or as power supplies fluctuate (supply noise). The research gap is the absence of principled, low-overhead methods for continuously computing and applying corrections that track these changes efficiently.

The research presented in this paper closes this gap directly: rather than treating physical drift as a nuisance to be corrected on a fixed schedule, it treats the drift itself as a measurable signal that a continuously running control system can act on in real time.

Research gap

No existing PNN system simultaneously achieves continuous adaptation to physical state changes, low energy overhead, hardware feasibility on low-cost substrates, and maintenance of primary task performance. This gap is the central challenge addressed by this research.

5.2 Specific research questions

The research presented in this paper is based on five specific research questions:

- How can physical-state variables (temperature, conductance drift, optical phase noise, supply voltage) be effectively encoded as informative signals for a learning-based adaptation controller?
- What is the optimal architecture for a meta-controller that takes physical state observations as input and output control signals to maintain PNN performance — balancing expressiveness with energy and latency constraints?
- How can the stability and convergence of a closed-loop adaptive system be guaranteed, given that the controlled physical system itself is non-stationary and exhibits complex dynamics?
- What are the fundamental trade-offs between adaptation speed, energy overhead and accuracy gains, and how do these trade-offs shift across different physical substrate technologies and deployment environments?
- Under what conditions can noise and physical non-ideality be constructively exploited — through stochastic resonance mechanisms — to improve rather than impair neural network performance?

6. Proposed system architecture

6.1 Architectural overview

The proposed system comprises two tightly coupled computational layers: a primary physical neural network (P-PNN) that implements the core inference or learning task, and a meta-controller (MC) that observes the physical state of the P-PNN and generates adaptive corrections. Figure 2 illustrates this dual-layer architecture and the information flows between components.

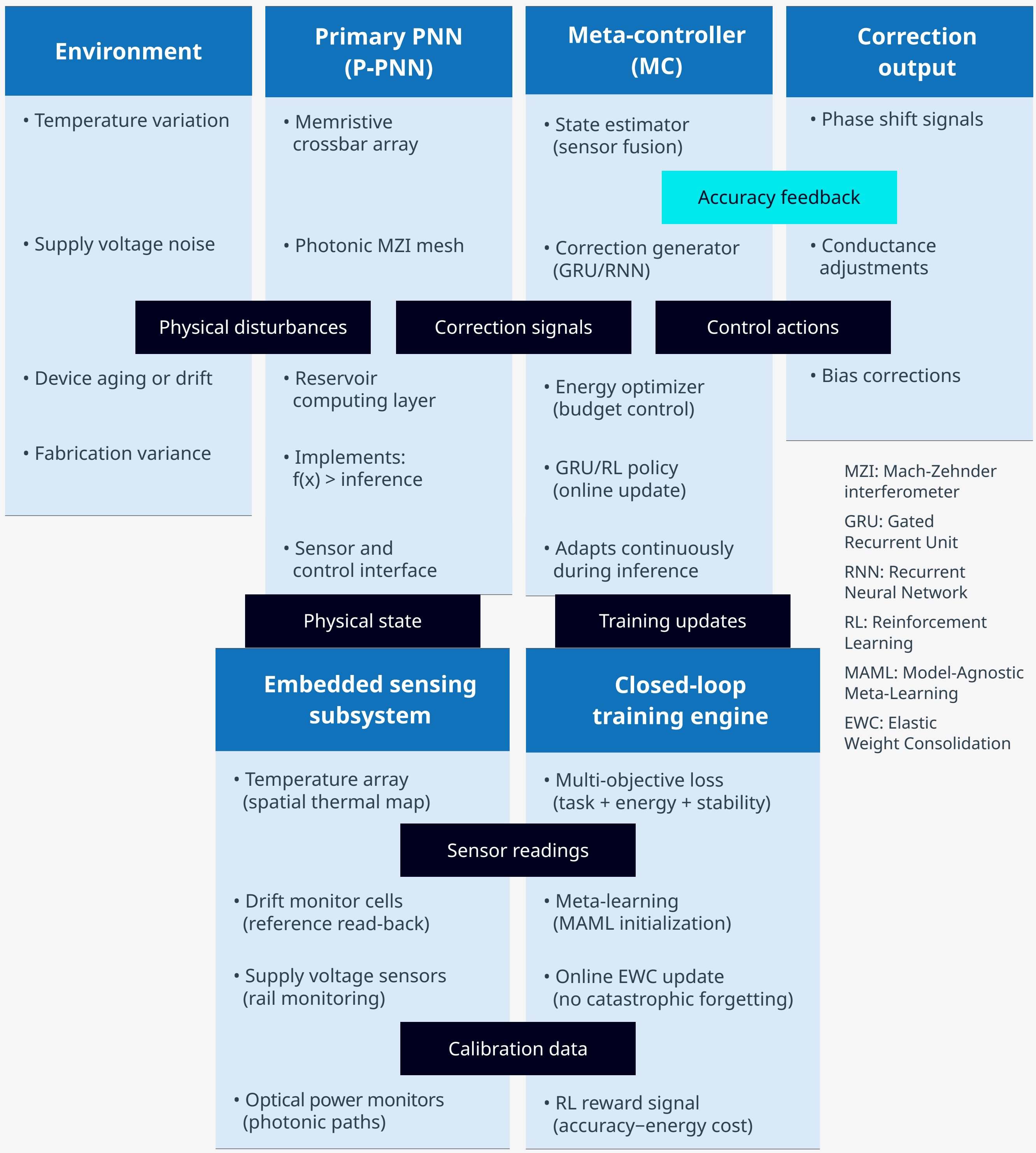
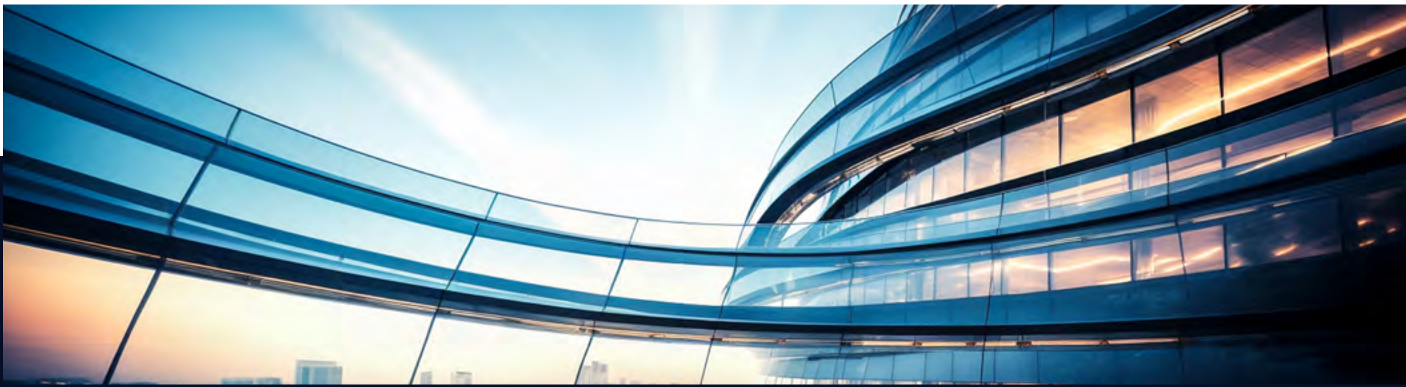


Figure 2: System architecture: The dual-layer adaptive PNN architecture, showing information flows between the primary PNN, embedded sensing subsystem, meta-controller and closed-loop training engine



Architectural key insight

The meta-controller does not need to model the full complexity of the physical substrate. It only needs to learn a mapping from observable sensor signals to effective control actions — a much lower-dimensional problem that can be solved with a lightweight, energy-efficient network.

6.2 Primary physical neural network layer

6.2.1 Memristive crossbar implementation

In the memristive implementation, the P-PNN is realized as one or more crossbar arrays of two-terminal memristive devices arranged at the intersections of horizontal wordlines and vertical bitlines. Analog weight values are stored as device conductance states, and matrix-vector multiplication is performed by applying input voltages to wordlines and reading out current sums on bitlines. Energy efficiencies in the range of 10 to 100 TOPS/W have been reported for representative architectures (Yao et al. 2020).

Primary non-idealities:

- Conductance drift over time
- Write noise (variability in the conductance achieved by a write pulse)
- Read noise (cycle-to-cycle variation in resistance measurement)
- Device-to-device variability from nanoscale fabrication differences

These non-idealities are quantified by the embedded sensing subsystem and fed to the meta-controller as a continuous state signal.

6.2.2 Photonic circuit implementation

In the photonic implementation, the P-PNN is realized as a programmable unitary matrix using a mesh of Mach–Zehnder interferometers (MZIs), each controlled by a thermal phase-shifter. The MZI mesh can implement any unitary transformation through the Clements or Reck decomposition (Clements et al. 2016; Miller 2015). Activation functions are implemented in electronics following optical-to-electronic conversion, enabling hybrid opto-electronic network architectures.

Primary non-idealities:

- Thermal phase drift (the most significant challenge), where ambient temperature changes produce phase errors in each phase-shifter that accumulate across the matrix
- Fabrication imperfections causing MZI splitting ratio deviations
- Optical loss variations
- Detector shot noise at the readout stage

6.3 Embedded sensing subsystem

Effective adaptation requires accurate and timely information about the physical state of the P-PNN. The embedded sensing subsystem is designed to provide this information at minimum energy and area overhead. The total sensing overhead is estimated at less than 5% of the inference energy budget for representative architectures.

Sensor type	Measurement target	Key purpose	Estimated overhead
Temperature array	Spatial chip temperature	Track thermal phase drift; conductance temperature coefficients	10–50 μW per sensor
Drift monitor cells	Reference device conductance/phase	Quantify drift rate without disrupting operational array	Negligible (periodic reads)
Supply voltage sensors	Power rail voltage	Detect supply noise and slow voltage drift	< 5 μW
Optical power monitors	Propagating optical power	Detect optical loss changes and coupling drift	1–10 μW
Activation monitors	Intermediate layer distributions	Indirect signal of accumulated weight error	< 1% inference overhead

Table 2: Sensor types for the embedded sensing subsystem

6.4 Meta-controller architecture

The meta-controller is the central algorithmic component. Its design is governed by three competing requirements:

- Expressive power to capture complex nonlinear relationships between sensor signals and optimal corrections
- Computational efficiency for real-time operation
- Sample efficiency to adapt rapidly from limited observations

The proposed meta-controller is a small recurrent neural network — specifically, a gated recurrent unit (GRU) with 32–128 hidden units — implemented on a low-power digital microcontroller or custom mixed-signal circuit. The recurrent architecture allows the meta-controller to integrate sensor information over time, which is essential for tracking slow drift processes that are not observable in a single timestep.

The meta-controller outputs a correction vector that parameterizes one of several correction mechanisms depending on the physical substrate:

- **For memristive arrays:** incremental conductance adjustments to high-importance weights
- **For photonic systems:** phase-shifter voltage corrections
- **For both:** scaling factors for analog bias currents compensating global drift

Selective correction strategy

Rather than computing corrections for all weights — which would require proportional write energy — the meta-controller prioritizes corrections to the most influential weights identified through sensitivity analysis. This reduces write energy by a factor of 10 to 100 while preserving most accuracy recovery.

7. Methodology and training framework

7.1 Closed-loop training philosophy

The central methodological innovation is a shift from the conventional linear deployment pipeline — Train ▶ Deploy ▶ Calibrate — to a closed-loop framework in which training is conducted in continuous interaction with the physical hardware or a high-fidelity simulation thereof. This closed-loop paradigm acknowledges that the physical substrate is an integral part of the computational system, requiring joint optimization. Figure 3 illustrates the complete training and adaptation flow.

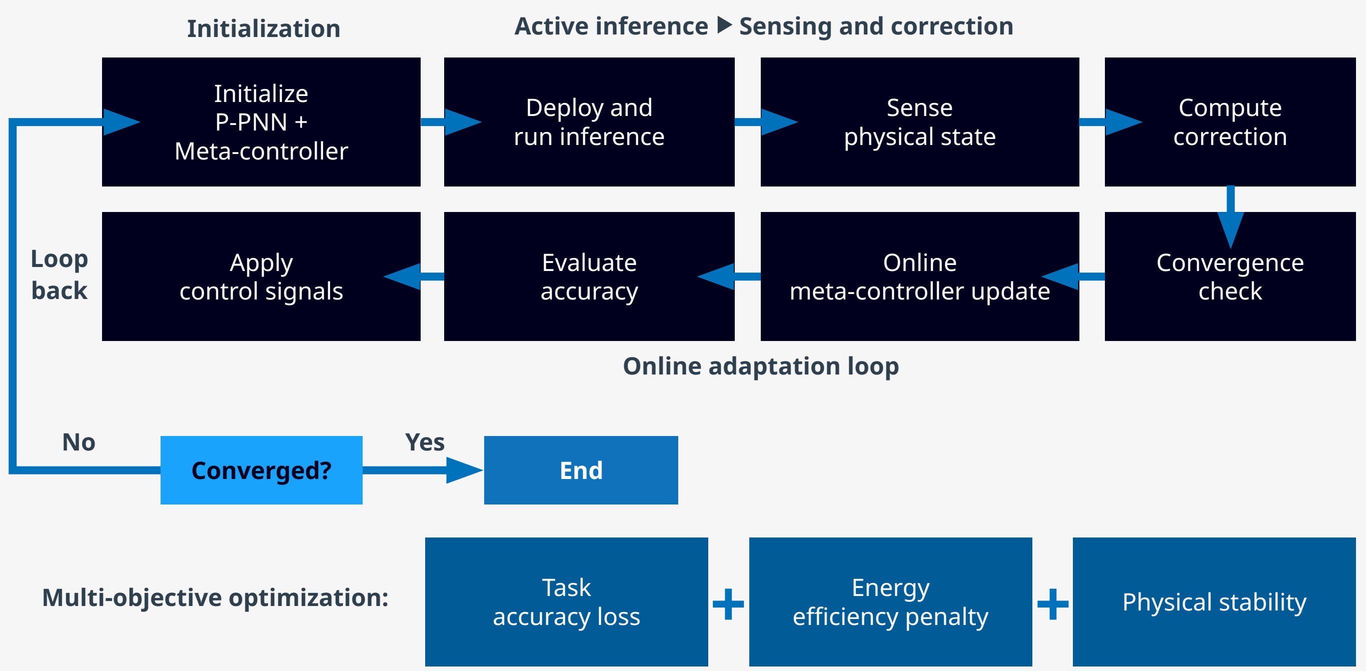


Figure 3: Closed-loop training framework: Complete process flow from initialization through active inference, sensing, correction and online adaptation, including the multi-objective optimization components

7.2 Multi-objective optimization

Training is guided by three competing objectives that reflect the real-world requirements of deployed PNN systems:

- **Task accuracy**
  The primary measure of whether PNN performs its intended function correctly, evaluated on the target dataset. This is the dominant constraint — all other objectives are secondary to maintaining useful task performance.
- **Energy efficiency**
  A penalty on the total number and magnitude of corrective write operations performed per inference, and on sensing subsystem power. This incentivizes the MC to achieve accuracy recovery with minimal physical intervention.
- **Physical stability**
  A regularization term that discourages large or rapid changes in the physical parameter space, preventing oscillatory or unstable control behavior. The weighting of this term is increased for long-lifetime deployment scenarios where conservative control behavior is preferred.

The relative weighting of these objectives is a design parameter that encodes deployment priorities. Extreme energy-constrained deployments (IoT nodes, implantable devices) prioritize energy efficiency; long-lifetime deployments prioritize stability; accuracy is always maintained as a primary constraint.

7.3 Meta-learning strategy

The meta-controller is trained using a meta-learning approach that prepares it to adapt rapidly to novel physical states it has not encountered during training. Drawing on the MAML framework of Finn et al. (2017), the meta-controller's initial parameters are optimized so that a small number of gradient steps — computed on sensor-correction pairs observed during deployment — bring the controller to a state that achieves effective correction on the current physical substrate state.

The meta-training procedure exposes the meta-controller to a distribution of physical perturbation scenarios, including temperature step changes, gradual drift trajectories, sudden supply-voltage drops and combined perturbation events. This meta-learning approach provides two key advantages: generalization to physical perturbation scenarios not seen during training, and the ability to adapt its own parameters during deployment as the substrate ages.

7.4 Online learning and deployment adaptation

During deployment, the meta-controller operates in an online learning mode, continuously updating its parameters from incoming sensor-accuracy feedback signals. Accuracy feedback is obtained through periodic evaluation on a small held-out calibration set — as few as 100 to 200 reference examples for classification tasks — evaluated at negligible cost relative to primary task throughput.

To prevent catastrophic forgetting of adaptation strategies learned during meta-training, the online learning procedure employs a modified elastic weight consolidation (EWC) regularizer (Kirkpatrick et al. 2017) that penalizes changes to parameters identified as important for the meta-learned prior. This ensures the meta-controller remains capable of rapid adaptation to common perturbation types while still tracking the specific drift trajectory of the deployed substrate.

7.5 Reinforcement learning for control policy optimization

Alongside gradient-based meta-learning, the research explores reinforcement learning as an alternative formulation for meta-controller optimization. In the reinforcement-learning formulation, the meta-controller is an agent that observes sensor readings and recent accuracy estimates, selects correction signals and receives rewards reflecting task accuracy improvement minus energy cost.

The proposed reinforcement-learning implementation uses a soft actor-critic (SAC) algorithm (Haarnoja et al. 2018), chosen for its sample efficiency and stable convergence in continuous action spaces. The policy and critic networks are implemented as compact multilayer perceptrons (2 to 3 layers; 64 to 28 units) to meet on-chip energy constraints. The reinforcement-learning approach is model-free, making it potentially more robust to unknown or poorly characterized physical effects than gradient-based approaches.

7.6 Stochastic resonance exploitation

The research specifically investigates the constructive use of noise through stochastic resonance mechanisms. In the stochastic-resonance regime, adding an optimal level of noise to input signals or intermediate activations can paradoxically improve classification performance by enhancing weak signal detection. The key practical question is whether the intrinsic device noise of a PNN operating at reduced power can be actively managed to remain within the beneficial stochastic-resonance regime, transforming noise from an adversary into a computational resource.

8. Experimental design and validation

8.1 Three-phase experimental program

The experimental program is organized into three sequential phases of increasing hardware fidelity, each building on the validated outputs of the phase before it. Figure 4 illustrates the sequence of phases with key milestones.

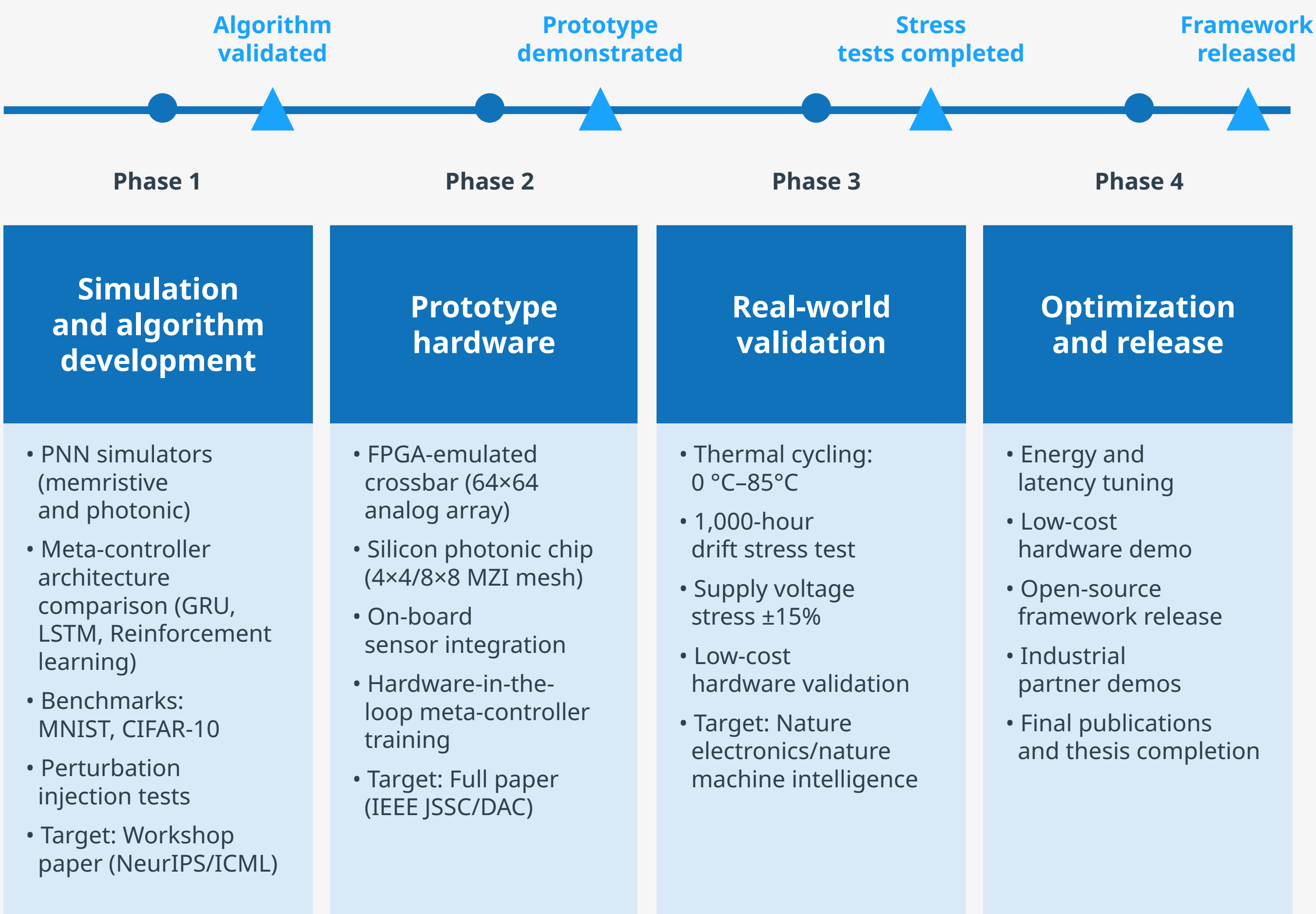


Figure 4: Research program phases: Sequential experimental phases from simulation and algorithm development through prototype hardware, real-world validation and final optimization and release.

DAC: Design Automation Conference

FPGA: Field-Programmable Gate Array

GRU: Gated Recurrent Unit

ICML: International Conference on Machine Learning

IEEE: Institute of Electrical and Electronics Engineers

JSSC: IEEE Journal of Solid-State Circuits

LSTM: Long Short-Term Memory

MNIST: Modified National Institute of Standards and Technology

MZI: Mach-Zehnder interferometer

NeurIPS: Neural Information Processing Systems conference

Phase 1: Simulation and algorithm development

Phase 1 establishes algorithmic foundations through high-fidelity simulation. PNN simulators for both memristive and photonic substrates are implemented, incorporating calibrated physical models of primary non-idealities: conductance drift, thermal phase noise and supply voltage fluctuation.

- Benchmark tasks: MNIST, CIFAR-10 and a temporal sequence classification task representative of edge inference applications
- Perturbation scenarios: Controlled injection of drift, thermal step changes (±20°C), supply noise and gradual device aging over simulated months of operation
- Meta-controller architectures compared: GRU, LSTM, convolutional temporal models and RL-based policies, evaluated on adaptation speed, steady-state accuracy recovery and computational cost
- Target output: Workshop paper submission to NeurIPS or ICML

Phase 2: Prototype hardware development

Phase 2 implements Phase 1 algorithms on prototype hardware, initially using an FPGA-based emulation of memristive crossbar arrays with programmable resistor networks, and subsequently on commercially available analog computing platforms.

- FPGA-emulated memristive platform: A 64×64 analog-emulation crossbar with programmable resistance and controlled noise injection, enabling rapid full-system prototyping without the lead times of memristive device fabrication
- Silicon photonic platform: A photonic neural network test chip fabricated in a standard 220nm silicon-on-insulator process via a multi-project wafer run, implementing a 4×4 or 8×8 MZI mesh
- Target output: Full research paper submission to IEEE JSSC or DAC

Phase 3: Real-world robustness testing

Phase 3 subjects the prototype to real-world stress conditions and evaluates long-term performance across four primary environmental challenges:

- Thermal cycling: Controlled temperature profiles from 0°C to 85°C (industrial temperature range), with both static and cycling profiles
- Long-term drift: Continuous 1,000-hour operation with hourly accuracy measurement, comparing adaptive and non-adaptive baselines
- Supply voltage stress: ±15% variation from nominal value, including transient voltage events simulating switching-converter noise
- Low-cost hardware generalization: Validation on deliberately commodity-grade hardware to assess democratized edge deployment potential

8.2 Evaluation metrics and success criteria

Table 3 sets out the metrics used to evaluate the performance of the proposed three-phase experimental program.

Metric	Definition	Success target
Task accuracy under perturbation	Primary task accuracy across all Phase 3 stress conditions	>95% of clean-state accuracy maintained
Calibration overhead reduction	Ratio of adaptive to baseline periodic recalibration energy	>80% energy overhead reduction
Adaptation latency	Time from perturbation onset to 90% accuracy recovery	<100 ms thermal; <10 ms voltage transients
Energy efficiency	Energy per inference and training step	>70% of ideal PNN efficiency advantage retained
Long-term stability	Standard deviation of accuracy over 1,000-hour operation	<2% accuracy standard deviation
Deployment lifetime extension	Ratio of time-to-5%-degradation (adaptive vs. static)	>3× lifetime extension demonstrated

Table 3: Evaluation metrics for the experimental program

9. Theoretical foundations

9.1 Stability analysis of closed-loop physical systems

The stability of the proposed closed-loop adaptive system is a central theoretical concern. The physical substrate evolves continuously according to internal dynamics influenced by environmental disturbances. The meta-controller implements a discrete-time control policy that reads sensor observations and applies corrections. The coupled dynamics of the physical system and the meta-controller create a hybrid discrete-continuous closed-loop system.

Stability analysis will employ both Lyapunov-based methods for deterministic drift dynamics and stochastic stability theory for noise-driven components. The key result sought is a characterization of the conditions under which the closed-loop system is input-to-state stable with respect to environmental disturbances — that is, bounded disturbances produce bounded deviations in task accuracy. Identifying these conditions will provide rigorous guarantees for deployed systems and inform the design of the stability term in the training objective.

9.2 Information-theoretic analysis of sensing

The choice of which physical quantities to sense, and at which sampling frequency, directly impacts both the effectiveness of adaptation and the energy cost of sensing. Information-theoretic analysis provides a principled framework for this design problem: quantifying how much information about correctable error is provided by the sensor readings.

This analysis guides sensor placement and selection, identifying the minimum sensor suite that provides sufficient information for effective meta-controller operation. It also informs the design of sensor fusion algorithms that optimally combine multiple sensor signals into a compact state estimate, reducing the dimensionality of the meta-controller input without losing predictive information.

9.3 Complexity bounds for adaptive correction

From a learning theory perspective, the meta-controller faces a sample complexity challenge: How many sensor-accuracy observations are required to learn a correction policy that achieves a given accuracy recovery level? This has practical implications for the minimum calibration dataset size and the frequency of online updates.

A key theoretical prediction is that the proposed adaptive framework achieves polynomial sample complexity in the number of drift dimensions, as compared to the exponential complexity that would be required if the full weight correction had to be learned without the physical model prior embedded in the meta-controller architecture. This provides a theoretical justification for the meta-learning approach and a quantitative limit on deployment overhead.

10. Expected contributions

This research shifts physical neural networks from being statically calibrated systems to having self-adaptive, physically-aware intelligence that learns to operate reliably in imperfect real-world conditions — transforming hardware non-ideality from a deployment barrier into a managed, and potentially exploitable, feature of computation.

Here, we explore the advancements and community contributions made by this research into areas like the theory and algorithms behind the PNN, along with the systems and the required hardware.

10.1 Theoretical contributions

- A formal framework for learning in non-stationary physical systems, extending existing online learning theory to account for the structured dynamics of physical substrates
- Stability guarantees for closed-loop physical neural networks — the first rigorous characterization of stability conditions for meta-controlled PNNs under realistic physical drift dynamics
- Information-theoretic bounds on the minimum sensing requirements for effective adaptation, providing principled guidance for sensor system design and placement
- New understanding of the relationship between physical noise statistics and neural network learning dynamics, including conditions under which stochastic resonance provides a net computational benefit

10.2 Algorithmic contributions

- A novel closed-loop training algorithm that jointly optimizes the P-PNN task parameters and the meta-controller adaptation policy, incorporating multi-objective loss terms for performance, energy efficiency and physical stability
- A physics-informed meta-learning strategy that initializes the meta-controller in a state conducive to rapid adaptation, exploiting the structured nature of physical drift dynamics to achieve sample-efficient online learning
- An online learning procedure for continual meta-controller adaptation during deployment, combining gradient-based updates with EWC regularization to maintain adaptation capability across long-term deployment
- Reinforcement learning control policies for PNN adaptation, providing a model-free alternative robust to uncharacterized physical effects

10.3 Systems and hardware contributions

- A prototype self-calibrating PNN hardware system — demonstrated on both emulated memristive and silicon photonic platforms — implementing the full adaptive pipeline from physical sensing through meta-controlled correction
- A documented reduction in calibration overhead relative to state-of-the-art periodic recalibration, quantified by energy consumption, latency and accuracy metrics across standardized benchmark conditions
- An open-source simulation framework for research in adaptive physical neural networks, including calibrated physical models of non-idealities, meta-controller baselines and benchmark tasks
- A hardware design reference for low-cost self-calibrating PNN systems, providing circuit-level guidance for integrating the adaptive framework on resource-constrained platforms

11. Broader impact and societal implications

11.1 Edge AI market context

The adaptive PNN architecture is designed to operate effectively on imperfect, low-cost hardware that cannot meet the precision specifications of conventional AI accelerators — a critical prerequisite for extending AI's benefits to resource-constrained markets. Figures 5 and 6 provide market and environmental context for this research program.

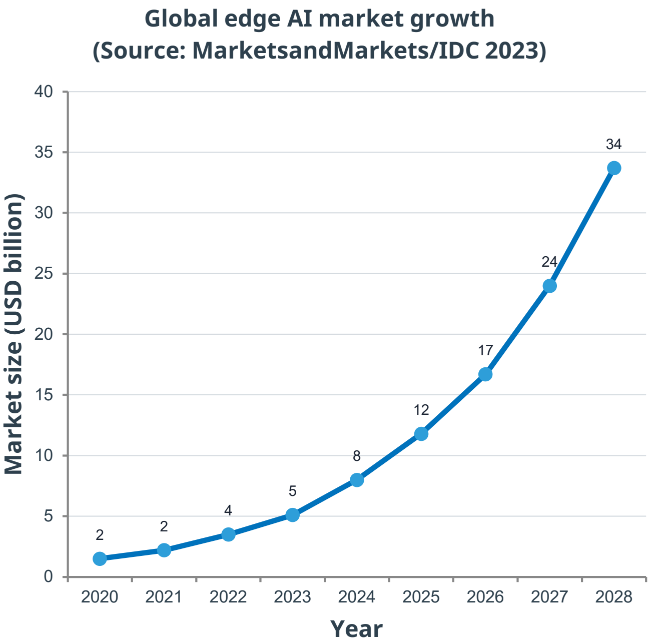


Figure 5: Market context

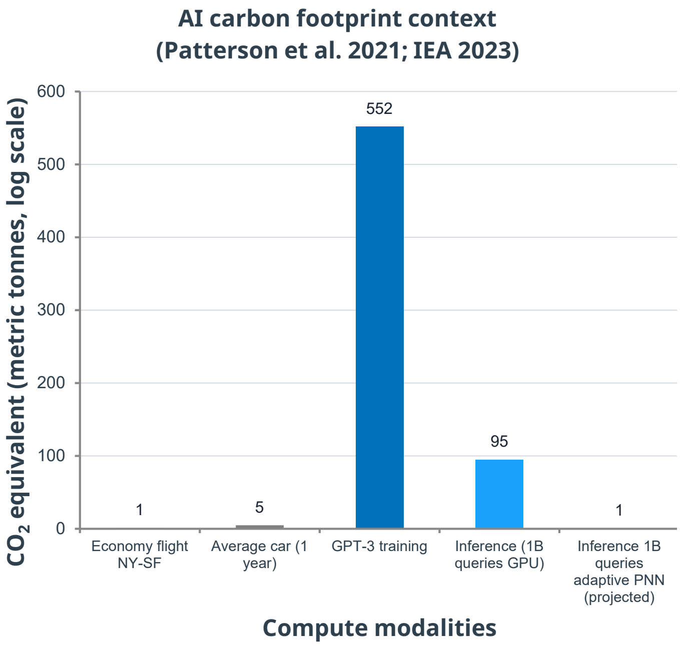


Figure 6: Environmental context

11.2 Democratizing AI at the edge

The most significant societal impact of this research is its potential to enable AI deployment on low-cost, resource-constrained devices — a prerequisite for extending AI's benefits to the approximately 4 billion people who currently lack reliable access to cloud-based AI infrastructure (Deng et al. 2020).

The proposed adaptive PNN architecture specifically targets operation on commodity analog hardware manufactured without expensive calibration processes. By enabling AI inference on lower-precision, lower-cost substrates, this deployment approach could reduce the hardware cost of AI inference by one to two orders of magnitude, placing it within reach of:

- Diagnostic AI for resource-limited healthcare settings in sub-Saharan Africa and South Asia
- Precision agriculture sensing for smallholder farms in Southeast Asia
- Low-power environmental monitoring for climate science in remote locations globally
- Low-cost industrial automation in developing markets

11.3 Environmental sustainability of AI

The energy consumption of AI infrastructure is a growing environmental concern. Patterson et al. (2021) estimate that training a large language model produces CO_2 emissions comparable to several car lifetimes. The energy efficiency advantage of PNNs — potentially 100× to 1,000× over digital GPU-based inference for matrix-heavy workloads — directly translates to reduced carbon emissions for AI inference at scale.

By extending the practical deployment lifetime of PNN hardware through adaptive calibration, the hardware refresh rate required to maintain performance is reduced, thereby reducing electronic waste. The combination of reduced inference energy and extended hardware lifetime positions adaptive PNNs as a key enabling technology for sustainable AI infrastructure.

Impact area	Mechanism	Projected outcome
Energy consumption	PNN inference energy 100× to 1,000× lower than GPU	Significant reduction in per-inference carbon footprint
Hardware lifetime	Adaptive calibration delays accuracy degradation	3x to 5× extension of practical deployment lifetime
Electronic waste	Longer hardware lifetime reduces refresh cycles	Proportional reduction in e-waste from AI hardware
AI accessibility	Lower hardware cost enables edge deployment	AI inference on commodity hardware in emerging markets
Industrial impact	New analog AI product classes enabled	Lower manufacturing cost via relaxed precision requirements

Table 4: How PNNs impact different areas to become an enabler for sustainable AI infrastructure

11.4 Hardware innovation and industrial impact

The research program's validation of self-calibrating PNN architectures on silicon photonic and memristive platforms will provide the semiconductor industry with new design principles for analog AI accelerator products. The demonstrated ability to operate effectively with lower-precision, lower-cost fabrication processes — relying on adaptive calibration rather than tight manufacturing tolerances — has direct implications for chip design methodologies, potentially enabling new classes of analog AI products at a substantially reduced manufacturing cost.

12. Real-world business applications and use cases

The architecture described in this paper is a research-based proposal, not a final product. However, the deployment scenarios it targets are already commercially significant today and are constrained precisely by the non-ideality problem this research addresses. This section outlines where self-calibrating, environment-adaptive PNNs would create concrete business value if the program succeeds.

12.1 Industrial IoT and predictive maintenance

Factory floors, oil and gas installations, and power infrastructure expose sensing hardware to vibration, electrical noise and wide thermal swings that are hostile to precisely calibrated analog electronics. An adaptive PNN embedded in a vibration or acoustic sensor could run continuous anomaly-detection inference at a fraction of the energy cost of a digital accelerator, without the periodic manual recalibration visits that make analog sensing unattractive to plant operators today. The business case is reduced unplanned downtime and lower cost per sensor node, at fleet scale.

12.2 Automotive and aerospace systems

Under-hood automotive electronics and satellite or drone payloads operate across some of the widest temperature and vibration ranges of any commercial electronics, and, in the case of spacecraft, cannot be physically serviced once deployed. A PNN that continuously self-calibrates would let manufacturers use lower-cost, lower-precision hardware in these environments while still meeting the accuracy and uptime guarantees safety-critical and business-critical systems require — directly addressing a barrier that today keeps analog accelerators out of automotive advanced driver assistance systems (ADAS) and satellite inference payloads.

12.3 Telecommunications and data center infrastructure

Photonic neural networks are a natural fit for outdoor 5G/6G base stations and for co-packaged optics inside data centers, where they could perform signal-processing or inference tasks at the speed of light with minimal added heat load. Both settings involve exactly the thermal and power-supply variability that this research targets. Operators would benefit from lower energy cost per inference at scale and from accelerators that hold their accuracy through routine equipment-heating cycles rather than degrading between maintenance windows.

12.4 Healthcare and wearable devices

Continuous health-monitoring wearables and implantable devices need always-on inference under strict battery and thermal constraints, sustained over months or years without clinical recalibration. A self-calibrating PNN would let device-makers extend battery life and operating lifetime simultaneously — a combination that is difficult to achieve with either digital accelerators (too power-hungry) or statically calibrated analog accelerators (too prone to drift) today.

12.5 Low-cost AI access in emerging markets

As discussed above, adaptive correction is what makes commodity-grade, loose-tolerance hardware usable for AI inference in the first place. This directly enables lower-cost AI-capable devices for agricultural monitoring, off-grid infrastructure, and rural connectivity in markets where uncontrolled ambient temperature and inconsistent power quality would otherwise rule out analog AI accelerators entirely, and where the cost of digital alternatives is prohibitive.

Sector	Deployment challenge today	Business value of adaptive PNNs
Industrial IoT	Vibration; electrical noise; manual recalibration cost	Lower downtime; lower cost per sensor node at fleet scale
Automotive and aerospace	Extreme temperature range; no post-deployment servicing	Lower-cost hardware qualifies for safety/ business-critical use
Telecom and data centers	Outdoor/rack thermal cycling; energy cost at scale	Lower energy cost per inference; sustained accuracy between maintenance windows
Healthcare wearables	Strict battery/thermal limits; no clinical recalibration	Longer battery life and operating lifetime together
Emerging markets	Uncontrolled ambient conditions; cost of digital alternatives	Commodity-grade hardware becomes commercially viable for AI

Table 5: Industrial sector wise challenges for ordinary PNNs and how adaptive PNNs help

These use cases share a common thread: each involves an environment the underlying physical substrate was not originally designed to tolerate. The research program's central assumption is that solving the general problem of physical drift once, at the architecture level, unlocks all of these markets simultaneously — rather than requiring a bespoke calibration solution engineered separately for each one.

12.6 Risks, challenges, and mitigation strategies

While there is a lot of potential that adaptive PNNs have in the future development of sustainable and efficient AI workloads, this novel technique comes with its own shares of risks and challenges. Table 6 lists all the challenges identified during this research, and the strategies to mitigate them.

Risk	Description	Mitigation strategy
Closed-loop instability	Coupled dynamics may exhibit oscillatory or divergent behavior under strong perturbations	Lyapunov stability analysis informs training regularization; conservative safety bounds on correction magnitude enforced at deployment
Sensor noise masking adaptation signal	Noise in the sensing subsystem may corrupt the physical state signal, caus-ing incorrect corrections	Kalman filtering and sensor fusion algorithms; redundant sensor modalities provide cross-validation of state estimates
Hardware-software codesign complexity	Tight integration of sensing, meta-controller and physical substrate creates risk of integration failures	Phased hardware approach with FPGA emulation before custom silicon; modular interface design isolates components
Meta-controller energy overhead	If the meta-controller energy footprint is disproportionate to accuracy gains, net system efficiency advantage is reduced	Systematic meta-controller architecture scaling study in Phase 1; hardware-efficient GRU implementations targeted for Phase 2
Limited physical hardware availability	Silicon photonic or memristive fabrication may be constrained by lead times, cost or yield	FPGA emulation provides a parallel development path; multi-project wafer runs reduce cost and lead time for photonic chips
Substrate generalization	Adaptation strategies optimized for one substrate may not transfer to another	Systematic cross-substrate evaluation in Phase 1 simulation; transfer learning exper-iments between substrate models

Table 6: Risks, challenges and mitigation strategies for adaptive PNNs

13. Performance benchmarks and supporting data

Having explored the theoretical and conceptual benefits in the working of adaptive PNNs, we now analyze available experimental data from these physical objects to understand more about their real-life accuracy, sustainability and efficiency. All data used here are sourced from peer-reviewed publications.

13.1 Energy-efficiency context

Figure 7 and 8 contextualize the energy-efficiency advantages of PNN substrates against conventional digital accelerators and illustrate the accuracy degradation trajectory that this research program aims to address.

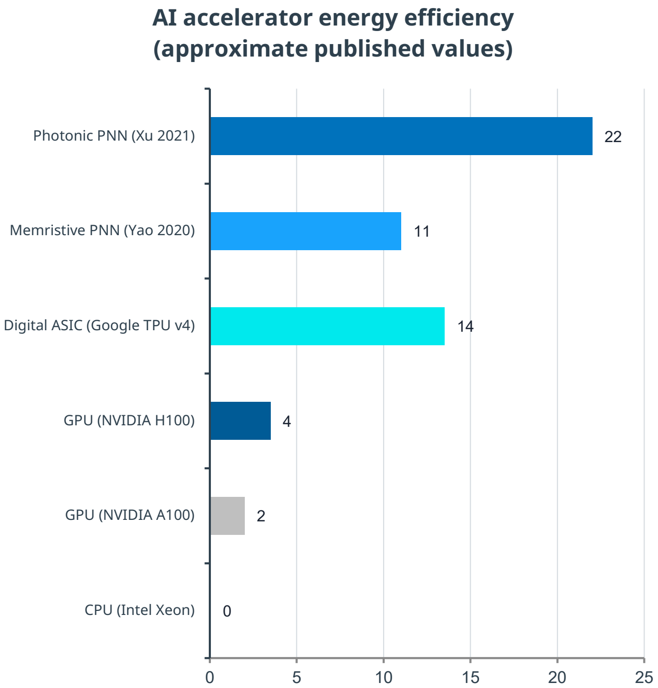


Figure 7: Energy context

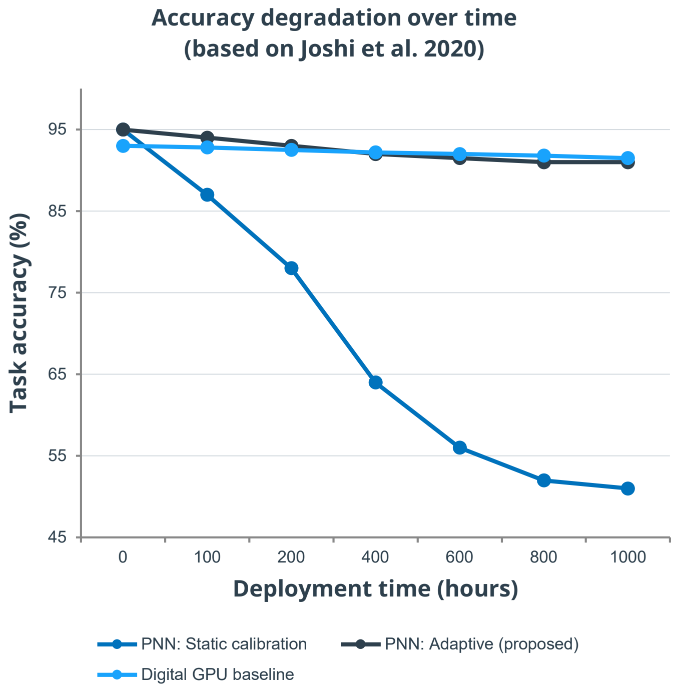


Figure 8: Accuracy context

13.2 Published performance data

Table 7 presents the publicly available performance data on adaptive PNNs from four major open source platforms.

Platform	Published efficiency (TOPS/W)	Task accuracy (benchmark)	Drift sensitivity	Source
NVIDIA H100 GPU	Photonic	Temperature changes shift phase-shifter operating point	Accumulated matrix error	NVIDIA 2023 Datasheet
Google tensor processing unit (TPU) v4	Memristive	Time-dependent change in programmed resistance	Weight value corruption	Google 2023
Memristive CNN (Yao 2020)	Memristive	Variability in conductance achieved by write pulse	Weight precision loss	Yao et al. 2020
Photonic Accelerator (Xu 2021)	Memristive	Cycle-to-cycle resistance measurement variation	Inference variance	Xu et al. 2021
PCM Analogue Inference	All	Nanoscale manufacturing imperfections	Systematic offset errors	Joshi et al. 2020
Adaptive PNN (projected)	All	Power rail fluctuations	Operating point shifts	This research program

Table 7: Publicly available performance data of adaptive PNNs

Note: TOPS/W figures reflect representative operating conditions from cited publications. Direct cross-platform comparisons should account for task complexity, precision requirements and measurement methodology differences.

14. Conclusion

Physical neural networks stand at an inflection point. The fundamental physics of their operation offers a route to AI-inference energy efficiencies that are unreachable by digital silicon CMOS — a route that could dramatically expand the reach and sustainability of AI deployment.

Yet this potential has been consistently undermined by the challenge of physical non-ideality: the gap between the ideal computational model of a PNN and the messy, drifting, temperature-sensitive reality of physical hardware in deployment.

The research program presented in this paper addresses this challenge through a conceptually elegant but technically deep insight: The physical state of a PNN is not merely an obstacle to be managed — it is an observable, learnable signal that, with the right architecture and algorithms, can be continuously monitored and compensated.

By integrating embedded sensing, meta-learning control and closed-loop training into a unified adaptive architecture, the proposed system transforms PNN deployment from a one-time calibration event into an ongoing adaptive process that maintains accuracy and efficiency throughout the entire lifetime of the device.

The three-phase experimental program provides a credible and rigorous path to demonstrating these claims. The theoretical foundations in stability analysis, information theory and learning theory underpin the empirical program with principled guarantees. The broader impact — in democratizing AI access, reducing AI's energy footprint and enabling new classes of low-cost hardware — provides compelling motivation for research investment, with concrete commercial pathways spanning industrial IoT, automotive and aerospace, telecommunications, healthcare and emerging-market deployment.

Program vision

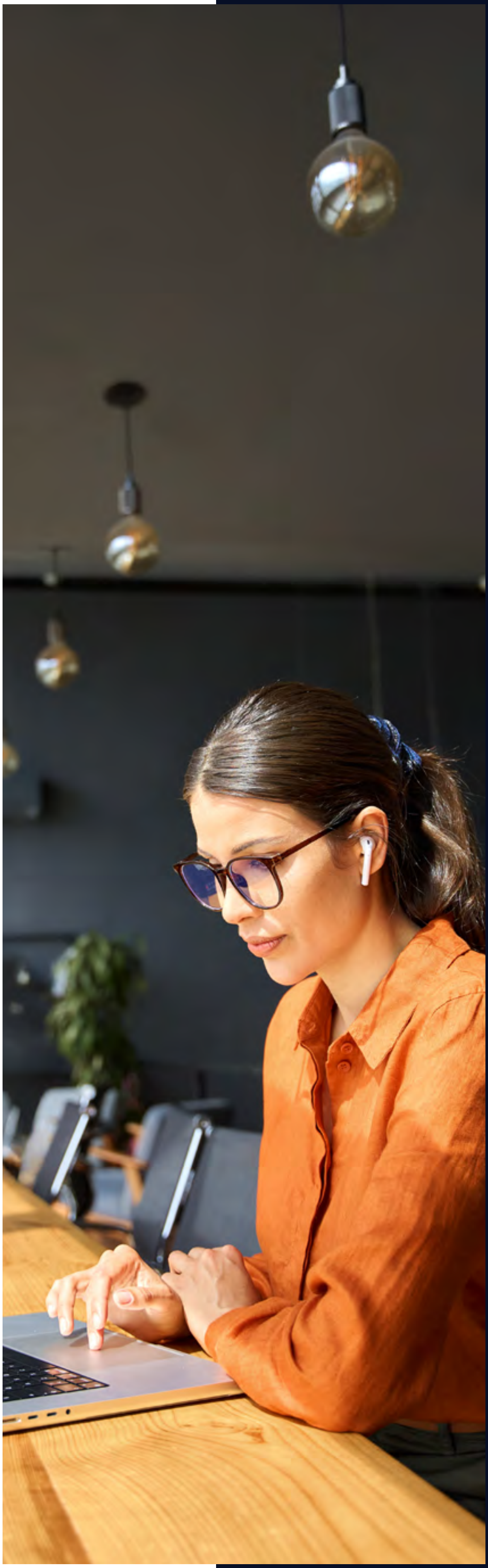
Self-calibrating, environment-adaptive PNNs represent the next necessary step in the maturation of physical computing from laboratory demonstration to real-world deployment. Success will establish the algorithmic, architectural and hardware foundations for a new generation of robust, efficient and democratically accessible AI systems.

15. Bibliography

The sources cited in and used to inform the contents of this paper are peer-reviewed publications or official technical documents from the named institutions:

- Ambrogio, Stefano, Pritish Narayanan, Hsinyu Tsai, Robert M. Shelby, Irem Boybat, Carmelo di Nolfo, Severin Sidler, Massimo Giordano, Martina Bodini, Nathan C. P. Farinha, et al. 2018. “Equivalent-Accuracy Accelerated Neural-Network Training Using Analogue Memory.” Nature 558 (7708): 60–67.
- Cai, Fuxi, Suhas Kumar, Thomas Van Vaerenbergh, Xia Sheng, Rui Liu, Can Li, Zhan Liu, Martin Foltin, Shimeng Yu, Qiangfei Xia, et al. 2020. “Power-Efficient Combinatorial Optimization Using Intrinsic Noise in Memristor Hopfield Neural Networks.” Nature Electronics 3 (7): 409–418.
- Clements, William R., Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley. 2016. “Optimal Design for Universal Multiport Interferometers.” Optica 3 (12): 1460–1465.
- Deng, Shuiguang, Hailiang Zhao, Weijia Fang, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. 2020. “Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence.” IEEE Internet of Things Journal 7 (8): 7457–7469.
- Finn, Chelsea, Pieter Abbeel, and Sergey Levine. 2017. “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks.” In Proceedings of the 34th International Conference on Machine Learning, 1126–1135.
- Gu, Jiaqi, Zheng Zhao, Chenghao Feng, Wuxi Li, Ray T. Chen, and David Z. Pan. 2020. “FLOPS: Efficient On-Chip Learning for Optical Neural Networks.” In Advances in Neural Information Processing Systems, 28–40.
- Haarnoja, Tuomas, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. “Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor.” In Proceedings of the 35th International Conference on Machine Learning, 1861–1870.
- Hospedales, Timothy M., Antreas Antoniou, Paul Micaelli, and Amos Storkey. 2021. “Meta-Learning in Neural Networks: A Survey.” IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (9): 5149–5169.
- International Energy Agency. 2024. Electricity 2024: Analysis and Forecast to 2026. Paris: International Energy Agency.
- Joshi, Vinay, Manuel Le Gallo, Simon Haefeli, Irem Boybat, Sascha R. Nandakumar, Christophe Piveteau, Martino Dazzi, Bipin Rajendran, Abu Sebastian, and Evangelos Eleftheriou. 2020. “Accurate Deep Neural Network Inference Using Computational Phase-Change Memory.” Nature Communications 11 (1): 2473.
- Kirkpatrick, James, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. 2017. “Overcoming Catastrophic Forgetting in Neural Networks.” Proceedings of the National Academy of Sciences 114 (13): 3521–3526.
- Larger, Laurent, Antonio Baylón-Fuentes, Romain Martinenghi, Vladimir S. Udaltsov, Yanne K. Chembo, and Maxime Jacquot. 2017. “High-Speed Photonic Reservoir Computing Using a Time-Delay-Based Architecture: Million Words per Second Classification.” Physical Review X 7 (1): 011015.
- MarketsandMarkets. 2023. Edge AI Software Market—Global Forecast to 2028. MarketsandMarkets Research.
- McDonnell, Mark D., and Derek Abbott. 2009. “What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology.” PLOS Computational Biology 5 (5): e1000348.
- Miller, David A. B. 2015. “Perfect Optics with Imperfect Components.” Optica 2 (8): 747–750.
- Nandakumar, Sascha R., Manuel Le Gallo, Irem Boybat, Bipin Rajendran, Abu Sebastian, and Evangelos Eleftheriou. 2018. “Mixed-Precision Architecture Based on Computational Memory for Training Deep Neural Networks.” In 2018 IEEE International Symposium on Circuits and Systems, 1–5.
- NVIDIA. 2023. NVIDIA H100 Tensor Core GPU Architecture White Paper. NVIDIA Corporation.
- Parisi, German I., Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. 2019. “Continual Lifelong Learning with Neural Networks: A Review.” Neural Networks 113: 54–71.
- Patterson, David, Joseph Gonzalez, Quoc V. Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. “Carbon Emissions and Large Neural Network Training.” arXiv. arXiv:2104.10350.

- Rasch, Malte J., Stefano Ambrogio, Pritish Narayanan, Hsinyu Tsai, Robert M. Shelby, Irem Boybat, Carmelo di Nolfo, Severin Sidler, Massimo Giordano, Martina Bodini, et al. 2019. "A Flexible and Fast Hardware Simulator for Brain-Inspired Computing Based on PCM Synapses." In 2019 IEEE International Symposium on Circuits and Systems.
- Sebastian, Abu, Manuel Le Gallo, Riduan Khaddam-Aljameh, and Evangelos Eleftheriou. 2020. "Memory Devices and Applications for In-Memory Computing." Nature Nanotechnology 15 (7): 529–544.
- Shastri, Bhavin J., Alexander N. Tait, Thomas Ferreira de Lima, Wolfram H. P. Pernice, Harish Bhaskaran, C. David Wright, and Paul R. Prucnal. 2021. "Photonics for Artificial Intelligence and Neuromorphic Computing." Nature Photonics 15 (2): 102–114.
- Shen, Yichen, Nicholas C. Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, and Marin Soljačić. 2017. "Deep Learning with Coherent Nanophotonic Circuits." Nature Photonics 11 (7): 441–446.
- Tanaka, Gouhei, Toshiyuki Yamane, Jean Benoît Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. 2019. "Recent Advances in Physical Reservoir Computing: A Review." Neural Networks 115: 100–123.
- Wright, Logan G., Tatsuhiro Onodera, Martin M. Stein, Tianyu Wang, Darren T. Schachter, Zoey Hu, and Peter L. McMahon. 2022. "Deep physical neural networks Trained with Backpropagation." Nature 601 (7894): 549–555.
- Xu, Xingyuan, Mengxi Tan, Bill Corcoran, Jiayang Wu, Andreas Boes, Thach G. Nguyen, Sai T. Chu, Brent E. Little, David G. Hicks, Roberto Morandotti, Arnan Mitchell, and David J. Moss. 2021. "11 TOPS Photonic Convolutional Accelerator for Optical Neural Networks." Nature 589 (7840): 44–51.
- Yao, Peng, Huaqiang Wu, Bin Gao, Jianshi Tang, Qingtian Zhang, Wenqiang Zhang, J. Joshua Yang, and He Qian. 2020. "Fully Hardware-Implemented Memristor Convolutional Neural Network." Nature 577 (7792): 641–646.



List of abbreviations

Abbreviation	Full term	Meaning
CMOS	Complementary Metal–Oxide–Semiconductor	A semiconductor technology used to build most modern integrated circuits, including microprocessors, memory chips, and image sensors.
PNN	Physical Neural Network	A neural network implemented directly in a physical substrate, such as photonic, memristive or analog hardware, where computation is performed by physical processes rather than digital logic.
MC	Meta-Controller	A lightweight adaptive controller that monitors the physical state of the PNN and generates corrective actions to compensate for drift, noise and environmental changes.
MAML	Model-Agnostic Meta-Learning	A meta-learning framework that trains models to adapt quickly to new tasks with minimal retraining.
EWC	Elastic Weight Consolidation	A continual learning technique that prevents catastrophic forgetting by protecting important parameters.
SAC	Soft Actor-critic	A reinforcement learning algorithm selected for efficient optimization of continuous control policies.
SR	Stochastic Resonance	A phenomenon where an optimal amount of noise improves signal detection or computational performance.
MZI	Mach–Zehnder Interferometer	A photonic component used to manipulate optical signals and implement matrix operations in photonic neural networks.
TOPS/W	Tera Operations Per Second/Per Watt	A standard metric used to measure computational energy efficiency.